Composition of Biochemical Networks using Domain Knowledge

Martin Goodfellow, John Wilson and Ela Hunt

Department of Computing and Information Sciences, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, UK
SystemsX.ch, ETH Zurich, Switzerland

EPSRC Engineering and Physical Sciences Research Council

Introduction

Composition and decomposition of graphs is an interesting and challenging area that has received little attention so far. Graph composition has applications in a variety of practical applications. In drug development, for instance, in order to understand possible drug interactions, one has to merge known networks and examine topological variants arising from such composition, or from graph splitting. The problem of nework composition has not been studied in depth, and has not received much attention. The aim of our research is to develop methods to perform network composition. The specific graphs we are interested in are biochemical networks. However, we hope to extend these methods to more general graphs, such as web and social networking graphs or large scale financial models.

Background

Graph matching is the initial step in both composition and decomposition. A lot of work has been performed in this area. In order to successfully cover the wide area of matching we created a taxonomy and identified a characteristic piece of work from each branch. We covered the following areas and work: Tree exact matching – XISS; approximate matching – ATreeGrep; General exact graph matching – VF2, GraphGrep; approximate matching – TALE, GraphFind; exact semi-structured data matching – Lore; exact molecule matching – Daylight; exact SBML matching - SBMLmerge; approximante metabolic network matching – Kahveci's algorithm; schema hybrid matching – S-Match; composite hybrid matching – QOM.

Apart from our work, the only existing composition software for SBML models is semanticSBML.

Graph Composition and Decomposition

Examples are shown in Figure 1.

SBMLCompose

SBMLCompose, performs composition of biochemical networks. Models are expressed in Systems Biology Markup Language (SBML) which is an XML based language for biochemical network modelling. SBMLCompose is part of BioNessie, which is a biochemical network simulation and analysis environment developed at the University of Glasgow.

We applied textual XML merging to SBML models.

One of the challenges here is the existence of arbitrary names and synonymy. To overcome this, we use synonym tables and the users are informed that biological entities must be given names expressing biological meaning. Mappings are stored to reduce comparison time further down the model.

Another challenge is how to handle inconsistencies/conflicts between identical components. The default is to issue a warning when a conflict is discovered. However, an interactive option is also included, to allow the user to interact with the software to resolve the conflict.

Test Results

SBMLCompose was developed in Java 1.6 and runs under Mac OS X, Linux and Windows.

semanticSBML, the only other SBML composition software, provides 17 test models. The size of these models ranges from 4 to 7 nodes and 0 to 3 edges. Each of these models was composed with every other model in the collection and the composition time recorded for both semanticSBML and SBMLCompose. The results of this experiment are shown in Figure 4. SBMLCompose is at least an order of magnitude faster than semanticSBML.

SBMLCompose Example

The models used in the composition are building blocks (Figure 3) which make up the Kholodenko model. This is a model of the MAPK/ERK signalling pathway. The Kholodenko biochemical network can be seen in Figure 2 along with its simulation. The oscillations present in the concentrations of the species over time are due to a feedback loop in the model.

The simulation for the model produced from the composition of building blocks 1 and 2 is shown in Figure 5. The simulation for the model of the three composed building blocks that was produced by the software is also shown, Figure 6. As can be seen from the graph the oscillations are in the same place but the colours of the lines differ compared with the simulation of the Kholodenko model. This is simply due to the fact that line colours are assigned to species in the order they are defined in the model. Therefore, as the composed model will have the species in a slightly different order, the line colours differ slightly. However, if the lines corresponding to each species from the graphs are compared it can be seen that the graphs are visually identical.

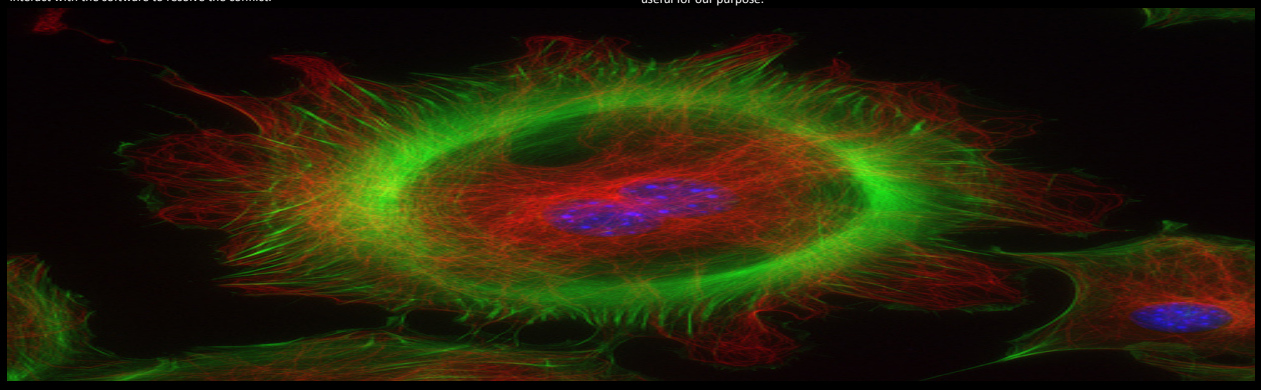
Beside visual comparison, other methods were used to evaluate test results:
textual comparison, residual sum of squares and the Monte Carlo Model Checker (MC2).

Conclusions and Current Work

We have demonstrated a new approach to graph merging which performs a simple textual merge.

Our work plan includes the following:
1. The development of new techniques to support efficient XML-based network merging and splitting
2. Using these methods with suitable domain knowledge representations so that they work for specific graphs in the biological domain

This has led us to research in the knowledge representation domain. We have looked at flat, hierarchical, network and XML based knowledge representations. We have identified schema and ontologies as the most useful for our purpose.



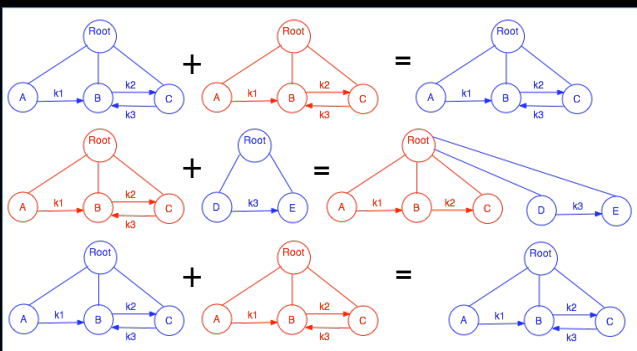
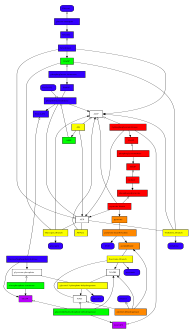


Figure 1



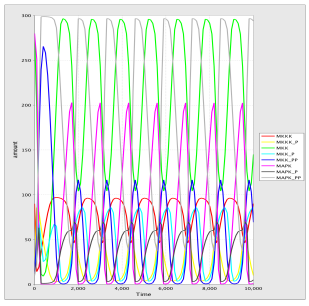


Figure 2

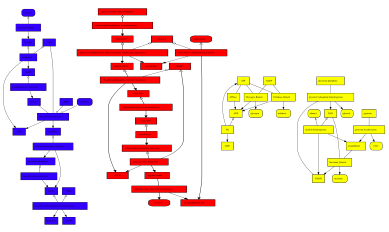


Figure 3

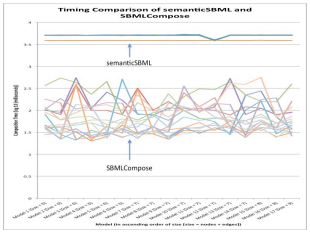


Figure 4

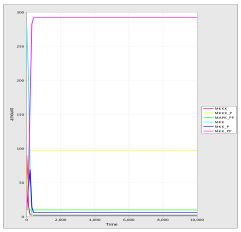


Figure 5

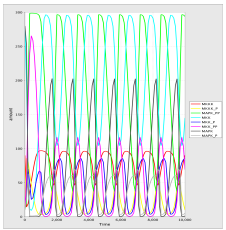


Figure 6